AN ELLIPSOIDAL BRANCH AND BOUND ALGORITHM FOR GLOBAL OPTIMIZATION *

WILLIAM W. HAGER† AND DZUNG T. PHAN‡

Abstract. A branch and bound algorithm is developed for global optimization. Branching in the algorithm is accomplished by subdividing the feasible set using ellipses. Lower bounds are obtained by replacing the concave part of the objective function by an affine underestimate. A ball approximation algorithm, obtained by generalizing of a scheme of Lin and Han, is used to solve the convex relaxation of the original problem. The ball approximation algorithm is compared to SEDUMI as well as to gradient projection algorithms using randomly generated test problems with a quadratic objective and ellipsoidal constraints.

Key words. global optimization, branch and bound, affine underestimation, convex relaxation, ball approximation, weakly convex

AMS subject classifications. 90C25, 90C26, 90C30, 90C45, 90C57

1. Introduction. In this paper we develop a branch and bound algorithm for the global optimization of the problem

$$\min \quad f(\mathbf{x}) \quad \text{subject to } \mathbf{x} \in \Omega, \tag{P}$$

where $\Omega \subset \mathbb{R}^n$ is a compact set and $f : \mathbb{R}^n \to \mathbb{R}$ is a weakly convex function [25]; that is, $f(\mathbf{x}) + \sigma\|\mathbf{x}\|^2$ is convex for some $\sigma \geq 0$. The algorithm starts with a known ellipsoid $\mathcal{E}$ containing Ω. The branching process in the branch and bound algorithm is based on successive ellipsoidal bisections of the original $\mathcal{E}$. A lower bound for the objective function value over an ellipse is obtained by writing f as the sum of a convex and a concave function and replacing the concave part by an affine underestimate. See [8, 13] for discussions concerning global optimization applications.

As a specific application of our global optimization algorithm, we consider problems with a quadratic objective function and with quadratic, ellipsoidal constraints. Global optimization algorithms for problems with quadratic objective function and quadratic constraints include those in [1, 18, 22]. In [22] Raber considers problems with nonconvex, quadratic constraints and with an n-simplex enclosing the feasible region. He develops a branch and bound algorithm based on a simplicial-subdivision of the feasible set and a linear programming relaxation over a simplex to estimate lower bounds. In a similar setting with box constraints, Linderoth [18] develops a branch and bound algorithm in which the the feasible region is subdivided using the Cartesian product of two-dimensional triangles and rectangles. Explicit formulae for the convex and concave envelops of bilinear functions over triangles and rectangles were derived. The algorithm of Le [1] focuses on problem with convex quadratic constraints; Lagrange duality is used to obtain lower bounds for the objective function, while ellipsoidal bisection is used to subdivide the feasible region.

The paper is organized as follows. In Section 2 we review the ellipsoidal bisection scheme of [1] which is used to subdivide the feasible region. Section 3 develops the

* June 28, 2008. This material is based upon work supported by the National Science Foundation under Grants 0619080 and 0620286.

†`hager@math.ufl.edu`, http://www.math.ufl.edu/∼hager, PO Box 118105, Department of Mathematics, University of Florida, Gainesville, FL 32611-8105. Phone (352) 392-0281. Fax (352) 392-8357.

‡`dphan@math.ufl.edu`, http://www.math.ufl.edu/∼dphan, PO Box 118105, Department of Mathematics, University of Florida, Gainesville, FL 32611-8105. Phone (352) 392-0281. Fax (352) 392-8357.

convex underestimator used to obtain a lower bound for the objective function. Since f is weakly convex, we can write it as the sum of a convex and concave functions:

$$f(\mathbf{x}) = \left(f(\mathbf{x}) + \sigma\|\mathbf{x}\|^2\right) + \left(-\sigma\|\mathbf{x}\|^2\right), \tag{1.1}$$

where $\sigma \geq 0$. A decomposition of this form is often called a DC (difference convex) decomposition (see [13]). For example, if f is a quadratic, then we could take

$$\sigma = -\min\{0, \lambda_1\},$$

where λ_1 is the smallest eigenvalue of the Hessian $\nabla^2 f$. The concave term $-\sigma\|\mathbf{x}\|^2$ in (1.1) is underestimated by an affine function ℓ which leads to a convex underestimate f_L of f given by

$$f_L(\mathbf{x}) = \left(f(\mathbf{x}) + \sigma\|\mathbf{x}\|^2\right) + \ell(\mathbf{x}). \tag{1.2}$$

We minimize f_L over the set $\mathcal{E} \cap \Omega$ to obtain a lower bound for the objective function on a subset of the feasible set. An upper bound for the optimal objective function value is obtained from the best feasible point produced when computing the lower bound, or from any local algorithm applied to this best feasible point. Note that weak convexity for a real-valued function is the analogue of hypomonotonicity for the derivative operator [7, 14, 21].

In Section 4 we discuss the phase one problem of finding a point in Ω which also lies in the ellipsoid $\mathcal{E}$. Section 5 gives the branch and bound algorithm and proves its convergence. Section 6 focuses on the special case where f and Ω are convex. The ball approximation algorithm of Lin and Han [16, 17] for projecting a point onto a convex set is generalized to replace the norm objective function by an arbitrary convex function. Numerical experiments, reported in Section 7, compare the ball approximation algorithm to SEDUMI 1.1 as well as to gradient projection algorithms. We also compare the branch and bound algorithm to a scheme of An [1] in which the lower bound is obtained by Lagrange duality.

Notation. Throughout the paper, $\|\cdot\|$ denotes the Euclidian norm. Given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $[\mathbf{x}, \mathbf{y}]$ is the line segment connecting $\mathbf{x}$ and $\mathbf{y}$:

$$[\mathbf{x}, \mathbf{y}] = \{(1-t)\mathbf{x} + t\mathbf{y} : 0 \leq t \leq 1\}.$$

The open line segment, which excludes the ends $\mathbf{x}$ and $\mathbf{y}$, is denoted $(\mathbf{x}, \mathbf{y})$. The interior of a set $\mathcal{S}$ is denoted int $\mathcal{S}$, while ri $\mathcal{S}$ is the relative interior. The gradient $\nabla f(\mathbf{x})$ is a row vector with

$$(\nabla f(\mathbf{x}))_i = \frac{\partial f(\mathbf{x})}{\partial x_i}.$$

The diameter of a set $\mathcal{S}$ is denoted $\delta(\mathcal{S})$:

$$\delta(\mathcal{S}) = \sup\ \{\|\mathbf{x} - \mathbf{y}\| : \mathbf{x},\ \ \mathbf{y} \in \mathcal{S}\}.$$

2. Ellipsoidal bisection. In this section, we give a brief overview of the ellipsoidal bisection scheme introduced by An [1]. This idea originates from the ellipsoid method for solving convex optimization problems by Shor, Nemirovski and Yudin [23, 27]. Consider an ellipsoid $\mathcal{E}$ with center $\mathbf{c}$ in the form

$$\mathcal{E} = \{\mathbf{x} \in \mathbb{R}^n : (\mathbf{x} - \mathbf{c})^{\mathsf{T}}\mathbf{B}^{-1}(\mathbf{x} - \mathbf{c}) \leq 1\}, \tag{2.1}$$

where $\mathbf{B}$ is a symmetric, positive definite matrix. Given a nonzero vector $\mathbf{v} \in \mathbb{R}^n$, the sets

$$H_- = \{\mathbf{x} \in \mathcal{E} : \mathbf{v}^\mathsf{T}\mathbf{x} \leq \mathbf{v}^\mathsf{T}\mathbf{c}\} \quad \text{and} \quad H_+ = \{\mathbf{x} \in \mathcal{E} : \mathbf{v}^\mathsf{T}\mathbf{x} \geq \mathbf{v}^\mathsf{T}\mathbf{c}\}$$

partition $\mathcal{E}$ into two sets of equal volume. The centers $\mathbf{c}_+$ and $\mathbf{c}_-$ and the matrix $\mathbf{B}_\pm$ of the ellipsoids $\mathcal{E}_\pm$ of minimum volume containing $H_\pm$ are given as follows:

$$\mathbf{c}_\pm = \mathbf{c} \pm \frac{\mathbf{d}}{n+1}, \quad \mathbf{B}_\pm = \frac{n^2}{n^2-1}\left(\mathbf{B} - \frac{2\mathbf{d}\mathbf{d}^\mathsf{T}}{n+1}\right), \quad \mathbf{d} = \frac{\mathbf{B}\mathbf{v}}{\sqrt{\mathbf{v}^\mathsf{T}\mathbf{B}\mathbf{v}}}.$$

As mentioned in [1], if the normal $\mathbf{v}$ always points along the major axis of $\mathcal{E}$, then a nested sequence of bisections shrinks to a point.

3. Bounding procedure. In this section, we obtain an affine underestimate ℓ for the concave function $-\|\mathbf{x}\|^2$ on the ellipsoid

$$\mathcal{E} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} - 2\mathbf{b}^\mathsf{T}\mathbf{x} \leq \rho\}, \tag{3.1}$$

where $\mathbf{A}$ is a symmetric, positive definite matrix, $\mathbf{b} \in \mathbb{R}^n$, and $\rho \in \mathbb{R}$. The set of affine underestimates for $-\|\mathbf{x}\|^2$ is given by

$$\mathcal{U} = \{\ell : \mathbb{R}^n \to \mathbb{R}, \ \ \ell \text{ is affine}, \ \ -\|\mathbf{x}\|^2 \geq \ell(\mathbf{x}) \text{ for all } \mathbf{x} \in \mathcal{E}\}. \tag{3.2}$$

The best underestimate is a solution of the problem

$$\min_{\ell \in \mathcal{U}} \ \max_{\mathbf{x} \in \mathcal{E}} \ -\left(\|\mathbf{x}\|^2 + \ell(\mathbf{x})\right). \tag{3.3}$$

THEOREM 3.1. *A solution of (3.3) is* $\ell^*(\mathbf{x}) = -2\mathbf{c}^\mathsf{T}\mathbf{x} + \gamma$*, where* $\mathbf{c} = \mathbf{A}^{-1}\mathbf{b}$ *is the center of the ellipsoid,* $\gamma = 2\mathbf{c}^\mathsf{T}\boldsymbol{\mu} - \|\boldsymbol{\mu}\|^2$*, and*

$$\boldsymbol{\mu} = \arg \max_{\mathbf{x} \in \mathcal{E}} \|\mathbf{x} - \mathbf{c}\|^2. \tag{3.4}$$

If $\delta(\mathcal{E})$ *is the diameter of* $\mathcal{E}$*, then*

$$\min_{\ell \in \mathcal{U}} \ \max_{\mathbf{x} \in \mathcal{E}} \ -\left(\|\mathbf{x}\|^2 + \ell(\mathbf{x})\right) = \frac{\delta(\mathcal{E})^2}{4}.$$

Proof. To begin, we will show that the minimization in (3.3) can be restricted to a compact set. Clearly, when carrying out the minimization in (3.3), we should restrict our attention to those ℓ which touch the function $h(\mathbf{x}) = -\|\mathbf{x}\|^2$ at some point in $\mathcal{E}$. Let $\mathbf{y} \in \mathcal{E}$ denote the point of contact. Since $h(\mathbf{x}) \geq \ell(\mathbf{x})$ and $h(\mathbf{y}) = \ell(\mathbf{y})$, a lower bound for the error $h(\mathbf{x}) - \ell(\mathbf{x})$ over $\mathbf{x} \in \mathcal{E}$ is

$$h(\mathbf{x}) - \ell(\mathbf{x}) \geq |\ell(\mathbf{x}) - \ell(\mathbf{y})| - |h(\mathbf{x}) - h(\mathbf{y})|.$$

If M is the difference between the maximum and minimum value of h over $\mathcal{E}$, then we have

$$h(\mathbf{x}) - \ell(\mathbf{x}) \geq |\ell(\mathbf{x}) - \ell(\mathbf{y})| - M. \tag{3.5}$$

An upper bound for the minimum in (3.3) is obtained by the function ℓ_0 which is constant on $\mathcal{E}$, with value equal to the minimum of $h(\mathbf{x})$ over $\mathbf{x} \in \mathcal{E}$. If $\mathbf{w}$ is a point where h attains its minimum over $\mathcal{E}$, then we have

$$\max_{\mathbf{x}\in\mathcal{E}} h(\mathbf{x}) - \ell_0(\mathbf{x}) = \max_{\mathbf{x}\in\mathcal{E}} h(\mathbf{x}) - h(\mathbf{w}) = M.$$

For $\mathbf{x} \in \mathcal{E}$, we have

$$h(\mathbf{x}) - \ell(\mathbf{x}) \le \max_{\mathbf{x}\in\mathcal{E}} h(\mathbf{x}) - \ell(\mathbf{x}) \le \max_{\mathbf{x}\in\mathcal{E}} h(\mathbf{x}) - \ell_0(\mathbf{x}) = M \tag{3.6}$$

when we restrict our attention to affine functions ℓ which achieve an objective function value in (3.3) which are at least as good as ℓ_0. Combining (3.5) and (3.6) gives

$$|\ell(\mathbf{x}) - \ell(\mathbf{y})| \le 2M \tag{3.7}$$

when ℓ achieves an objective function value in (3.3) which is at least as good as ℓ_0. Thus, when we carry out the minimization in (3.3), we should restrict to affine functions which touch h at some point $\mathbf{y} \in \mathcal{E}$ and with the change in ℓ across $\mathcal{E}$ satisfying the bound (3.7) for all $\mathbf{x} \in \mathcal{E}$. This tells us that the minimization in (3.3) can be restricted to a compact set, and that a minimizer must exist.

Suppose that ℓ attains the minimum in (3.3). Let $\mathbf{z}$ be a point in $\mathcal{E}$ where $h(\mathbf{x}) - \ell(\mathbf{x})$ achieves its maximum. A Taylor expansion around $\mathbf{x} = \mathbf{z}$ gives

$$h(\mathbf{x}) - \ell(\mathbf{x}) = h(\mathbf{z}) - \ell(\mathbf{z}) + (\nabla h(\mathbf{z}) - \nabla \ell)(\mathbf{x} - \mathbf{z}) - \|\mathbf{x} - \mathbf{z}\|^2 \tag{3.8}$$

since $h(\mathbf{x}) = -\|\mathbf{x}\|^2$. Since $\ell \in \mathcal{U}$, the set given in (3.2), we have $h(\mathbf{x}) - \ell(\mathbf{x}) \ge 0$ for all $\mathbf{x} \in \mathcal{E}$, so (3.8) yields

$$0 \le h(\mathbf{z}) - \ell(\mathbf{z}) + (\nabla h(\mathbf{z}) - \nabla \ell)(\mathbf{x} - \mathbf{z}) - \|\mathbf{x} - \mathbf{z}\|^2 \tag{3.9}$$

for all $\mathbf{x} \in \mathcal{E}$. By the first-order optimality conditions for $\mathbf{z}$, we have

$$(\nabla h(\mathbf{z}) - \nabla \ell)(\mathbf{x} - \mathbf{z}) \le 0$$

for all $\mathbf{x} \in \mathcal{E}$. It follows from (3.9) that

$$0 \le h(\mathbf{z}) - \ell(\mathbf{z}) - \|\mathbf{x} - \mathbf{z}\|^2,$$

or

$$h(\mathbf{z}) - \ell(\mathbf{z}) \ge \|\mathbf{x} - \mathbf{z}\|^2$$

for all $\mathbf{x} \in \mathcal{E}$. Since there exists $\mathbf{x} \in \mathcal{E}$ such that $\|\mathbf{x} - \mathbf{z}\| \ge \delta(\mathcal{E})/2$, we have

$$\max_{\mathbf{x}\in\mathcal{E}} h(\mathbf{x}) - \ell(\mathbf{x}) = h(\mathbf{z}) - \ell(\mathbf{z}) \ge \delta(\mathcal{E})^2/4. \tag{3.10}$$

We now observe that for the specific affine function ℓ^* given in the statement of the theorem, (3.10) becomes an equality, which implies the optimality of ℓ^* in (3.3). Expand in a Taylor series around $\mathbf{x} = \mathbf{c}$, where $\mathbf{c} = \mathbf{A}^{-1}\mathbf{b}$ is the center of the ellipsoid $\mathcal{E}$, to obtain

$$h(\mathbf{x}) = -\|\mathbf{c}\|^2 - 2\mathbf{c}^{\mathsf{T}}(\mathbf{x} - \mathbf{c}) - \|\mathbf{x} - \mathbf{c}\|^2 = -2\mathbf{c}^{\mathsf{T}}\mathbf{x} + \|\mathbf{c}\|^2 - \|\mathbf{x} - \mathbf{c}\|^2.$$

Hence, for ℓ^*, we have

$$h(\mathbf{x}) - \ell^*(\mathbf{x}) = \|\mathbf{c}\|^2 - \gamma - \|\mathbf{x}-\mathbf{c}\|^2 = \|\boldsymbol{\mu}-\mathbf{c}\|^2 - \|\mathbf{x}-\mathbf{c}\|^2$$
$$= \max_{\mathbf{y}\in\mathcal{E}} \|\mathbf{y}-\mathbf{c}\|^2 - \|\mathbf{x}-\mathbf{c}\|^2.$$

Clearly, $h(\mathbf{x}) - \ell^*(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathcal{E}$, and the maximum over $\mathbf{x} \in \mathcal{E}$ is attained at $\mathbf{x} = \mathbf{c}$. Moreover,

$$h(\mathbf{c}) - \ell^*(\mathbf{c}) = \max_{\mathbf{y}\in\mathcal{E}} \|\mathbf{y}-\mathbf{c}\|^2 = \delta(\mathcal{E})^2/4.$$

Consequently, (3.10) becomes an equality for $\ell = \ell^*$, which implies the optimality of ℓ^* in (3.3). ▯

To evaluate the best affine underestimate given by Theorem 3.1, we need to solve the optimization problem (3.4). This amounts to finding the major axis of the ellipsoid. The solution is

$$\boldsymbol{\mu} = \mathbf{c} + s\mathbf{y},$$

where $\mathbf{y}$ is a unit eigenvector of $\mathbf{A}$ associated with the smallest eigenvalue ϵ, and s is chosen so that $\boldsymbol{\mu}$ lies on the boundary of the $\mathcal{E}$. From the definition of $\mathcal{E}$, we obtain

$$s = \sqrt{(\mathbf{c}^{\mathsf{T}}\mathbf{A}\mathbf{c} + \rho)/\epsilon}.$$

We minimize the function f_L in (1.2) over $\mathcal{E} \cap \Omega$, with ℓ the best affine underestimate of $-\|\mathbf{x}\|^2$, to obtain a lower bound for the objective function over $\mathcal{E} \cap \Omega$. An upper bound for the optimal objective function value is obtained by starting any local optimization algorithm from the best iterate generated during the computation of the lower bound. For the numerical experiments reported later, the gradient projection algorithm [11] is the local optimization algorithm. Of course, by using a faster local algorithm, the overall speed of the global optimization algorithm will increase.

4. Phase one. In each step of the branch and bound algorithm for (P), we need to solve a problem of the form

$$\min \quad f(\mathbf{x}) \quad \text{subject to } \mathbf{x} \in \mathcal{E} \cap \Omega, \tag{4.1}$$

in the special case where f is convex (the function f_L in (1.2)) and $\mathcal{E}$ is an ellipsoid. In order to solve this problem, we often need to find a feasible point. One approach for finding a feasible point is to consider the minimization problem

$$\min \quad \mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x} - 2\mathbf{b}^{\mathsf{T}}\mathbf{x} \quad \text{subject to } \mathbf{x} \in \Omega, \tag{4.2}$$

where $\mathbf{A}$ and $\mathbf{b}$ are associated with the ellipsoid $\mathcal{E}$ in (3.1). Assuming we know a feasible point $\mathbf{x}_0 \in \Omega$, we could apply an optimization algorithm to (4.2). If the objective function value can be reduced below ρ, then we obtain a point in $\mathcal{E}$. If the optimal objective function value is strictly larger than ρ, then the problem (4.1) is infeasible.

If the set Ω is itself the intersection of ellipsoids, then the procedure we have just described could be used in a recursive fashion to determine a feasible point for either Ω or $\mathcal{E} \cap \Omega$, if it exists. In particular, suppose $\Omega = \cap_{j=1}^m \mathcal{E}_j$ is the intersection of m ellipsoids, where

$$\mathcal{E}_j = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}^{\mathsf{T}}\mathbf{A}_j\mathbf{x} - 2\mathbf{b}_j^{\mathsf{T}}\mathbf{x} \leq \rho_j\}.$$

A point $\mathbf{x}_1 \in \mathcal{E}_1$ is readily determined. Proceeding by induction, suppose that we have a point $\mathbf{x}_{k-1} \in \cap_{j=1}^{k-1}\mathcal{E}_j$. Any globally convergent iterative method is applied to the convex optimization problem

$$\min \ \mathbf{x}^\mathsf{T}\mathbf{A}_k\mathbf{x} - 2\mathbf{b}_k^\mathsf{T}\mathbf{x} \quad \text{subject to } \mathbf{x} \in \cap_{j=1}^{k-1}\mathcal{E}_j.$$

If the objective function value is reduced below ρ_k, then a feasible point in $\cap_{j=1}^{k}\mathcal{E}_j$ has been determined. Conversely, if the optimal objective function value is above ρ_k, then $\cap_{j=1}^{k}\mathcal{E}_j$ is empty.

5. Branch and bound algorithm. Our branch and bound algorithm is patterned after a general branch and bound algorithm, as appears in [13] for example. For any ellipse $\mathcal{E}$, define

$$M_L(\mathcal{E}) = \min \ \{f_L(\mathbf{x}) : \mathbf{x} \in \mathcal{E} \cap \Omega\}, \tag{5.1}$$

where f_L is the lower bound (1.2) corresponding to the best affine underestimate of $-\|\mathbf{x}\|^2$ on $\mathcal{E}$. We assume that an algorithm is available to solve the optimization problem (5.1).

Ellipsoidal branch and bound with linear underestimate (EBL)

1. Let $\mathcal{E}_0$ be an ellipsoid which contains Ω and set $\mathcal{S}_0 = \{\mathcal{E}_0\}$.
2. Evaluate $M_L(\mathcal{E}_0)$ and let $\mathbf{x}_0 \in \Omega$ denote the feasible point generated during the evaluation of $M_L(\mathcal{E}_0)$ with the smallest function value.
3. For $k = 0, 1, 2, \ldots$
   (a) Choose $\mathcal{E}_k \in \mathcal{S}_k$ such that $M_L(\mathcal{E}_k) = \min\{M_L(\mathcal{E}) : \mathcal{E} \in \mathcal{S}\}$. Bisect $\mathcal{E}_k \in \mathcal{S}_k$ with two ellipsoids denoted $\mathcal{E}_{k1}$ and $\mathcal{E}_{k2}$ (see Section 2). Evaluate $M_L(\mathcal{E}_{k1})$ and $M_L(\mathcal{E}_{k2})$.
   (b) Let $\mathbf{x}_{k+1}$ denote a feasible point associated with the smallest function value that has been generated up to this iteration and up to this step. Hence, if $\mathbf{y}_{k1}$ and $\mathbf{y}_{k2}$ are solutions to (5.1) associated with $\mathcal{E} = \mathcal{E}_{k1}$ and $\mathcal{E} = \mathcal{E}_{k2}$ respectively, then we have $f(\mathbf{x}_{k+1}) \le f(\mathbf{y}_{ki})$, $i = 1, 2$.
   (c) Set $\mathcal{S}_{k+1} = \{\mathcal{E} \in \mathcal{S}_k \cup \{\mathcal{E}_{k1}\} \cup \{\mathcal{E}_{k2}\} : M_L(\mathcal{E}) \le f(\mathbf{x}_{k+1}), \mathcal{E} \ne \mathcal{E}_k\}$

THEOREM 5.1. *Suppose that the following conditions hold:*
A1. The feasible set Ω is contained in some given ellipsoid $\mathcal{E}$, Ω is compact, and f is weakly convex over $\mathcal{E}$.
A2. A nested sequence of ellipsoidal bisections shrinks to a point (see Section 2).
Then every accumulation point of the sequence $\mathbf{x}_k$ is a solution of (P).

Proof. Let $\mathbf{y}$ denote any global minimizer for (P). We now show that for each k, there exists $\mathcal{E} \in \mathcal{S}_k$ with $\mathbf{y} \in \mathcal{E}$. Since $\Omega \subset \mathcal{E}_0$, $\mathbf{y} \in \mathcal{E}_0$. Proceeding by induction, suppose that for each j, $0 \le j \le k$, there exists an ellipsoid $\mathcal{F}_j \in \mathcal{S}_j$ with $\mathbf{y} \in \mathcal{F}_j$. We now wish to show that there exist $\mathcal{F}_{k+1} \in \mathcal{S}_{k+1}$ with $\mathbf{y} \in \mathcal{F}_{k+1}$. In Step 3c, $\mathcal{F}_k \in \mathcal{S}_k$ can only be deleted from $\mathcal{S}_{k+1}$ if $M_L(\mathcal{F}_k) > f(\mathbf{x}_{k+1})$ or $\mathcal{F}_k = \mathcal{E}_k$. The former case cannot occur since

$$M_L(\mathcal{F}_k) \le f(\mathbf{y}) \le f(\mathbf{x}_{k+1}),$$

due to the global optimality of $\mathbf{y}$. If $\mathcal{F}_k = \mathcal{E}_k$, then $\mathbf{y}$ lies in either $\mathcal{E}_{k1}$ or $\mathcal{E}_{k2}$. If $\mathbf{y} \in \mathcal{E}_{ki}$, then $\mathcal{E}_{ki} \in \mathcal{S}_{k+1}$ since

$$M_L(\mathcal{E}_{ki}) \le f(\mathbf{y}) \le f(\mathbf{x}_{k+1}).$$

Let $\mathbf{x}^*$ denote an accumulation point of the sequence $\mathbf{x}_k$. Since Ω is closed and $\mathbf{x}_k \in \Omega$ for each k, $\mathbf{x}^* \in \Omega$. By [25, Prop. 4.4], a weakly convex function is locally Lipschitz continuous. Hence, f is continuous on Ω and $f(\mathbf{x}_k)$ approaches $f(\mathbf{x}^*)$. If $\mathbf{x}^*$ is a solution of (P), then the proof is complete. Otherwise, $f(\mathbf{y}) < f(\mathbf{x}^*)$.

For each k, we have

$$\min \ \{M_L(\mathcal{E}) : \mathcal{E} \in \mathcal{S}_k\} \le M_L(\mathcal{F}_k) \le f(\mathbf{y}) < f(\mathbf{x}^*). \tag{5.2}$$

Let $\mathcal{G}_k$ denote an ellipsoid which achieves the minimum on the left side of (5.2) and let $\mathbf{y}_k$ denote a minimizer in (5.1) corresponding to $\mathcal{E} = \mathcal{G}_k$. The inequality (5.2) reduces to

$$f_L(\mathbf{y}_k) \le f(\mathbf{y}) < f(\mathbf{x}^*). \tag{5.3}$$

Since $\mathbf{y}$ minimizes f over Ω, (5.3) implies that

$$f_L(\mathbf{y}_k) \le f(\mathbf{y}) \le f(\mathbf{y}_k). \tag{5.4}$$

By Theorem 3.1,

$$f(\mathbf{y}_k) - f_L(\mathbf{y}_k) = -\sigma(\|\mathbf{y}_k\|^2 + \ell(\mathbf{y}_k)) \le \sigma\delta(\mathcal{G}_k)^2/4, \tag{5.5}$$

where ℓ is the best linear lower bound for the function $-\|\mathbf{x}\|^2$, and $\sigma \ge 0$ is the parameter associated with the convex/concave decomposition (1.1).

Each ellipsoid $\mathcal{E}_k$ corresponds to a vertex on the branch and bound tree associated with EBL. Choose the iteration numbers $k_1 < k_2 < \ldots$ so that they correspond to vertices along an infinite path on the branch and bound tree, starting from the root of the tree. By (A2), $\delta(\mathcal{G}_{k_i})$ tends to 0 as i tends to infinity. Hence, (5.5) implies that $|f(\mathbf{y}_{k_i}) - f_L(\mathbf{y}_{k_i})|$ tends to zero. Combining this with (5.3) and (5.4) shows that $f(\mathbf{y}_{k_i}) < f(\mathbf{x}^*)$ for i sufficiently large, which violates Step 3b and the fact that $f(\mathbf{x}_{k+1})$ is the smallest function value at step k and the smallest values monotonically approach $f(\mathbf{x}^*)$. □

Note that if for any k, $f(\mathbf{x}_k) = \min\{M_L(\mathcal{E}) : \mathcal{E} \in \mathcal{S}_k\}$, then $\mathbf{x}_k$ is a global minimizer.

6. Ball approximation algorithm for convex optimization. In this section we give an algorithm to solve (P) in the special case that f and Ω are convex. This algorithm, which is based on the successive approximation of the feasible set by balls, ties in nicely with the ellipsoidal-based branch and bound algorithm. The algorithm is a generalization of the ball approximation algorithm [17] of Lin and Han. The algorithm of Lin and Han deals with the special case where the objective function has the form $\|\mathbf{x}-\mathbf{a}\|^2$ and Ω is an intersection of ellipsoids. Lin generalizes this algorithm in [16] to treat convex constraints. The analysis in [16, 17] is tightly coupled to the norm objective function. In our further generalization of the Lin/Han algorithm, the norm objective function is replaced by an arbitrary convex functional f and an additional constraint set $\chi \subset \mathbb{R}^n$ is included, which might represent bound constraints for example. More precisely, we consider the problem

$$\min f(\mathbf{x}) \quad \text{subject to } \mathbf{x} \in \mathcal{F} := \{\mathbf{x} \in \chi : \mathbf{g}(\mathbf{x}) \le \mathbf{0}\}, \tag{C}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$, and the following conditions hold:

C1. f and $\mathbf{g}$ are convex and differentiable, χ is closed and convex, and $\mathcal{F}$ is compact.

C2. There exists $\bar{\mathbf{x}}$ in the relative interior of χ with $\mathbf{g}(\bar{\mathbf{x}}) < \mathbf{0}$.
C3. There exists $\gamma > 0$ such that $\|\nabla g_i(\mathbf{x})\| \geq \gamma$ when $g_i(\mathbf{x}) = 0$ for some $i \in [1, m]$ and $\mathbf{x} \in \chi$.

The condition C2 is referred to as the Slater condition.

We will give a new analysis which handles this more general convex problem (C). In each iteration of Lin's algorithm in [16], the convex constraints are approximated by ball constraints. Let $h : \mathbb{R}^n \to \mathbb{R}$ be a convex, differentiable function which defines a convex, nonempty set

$$\mathcal{H} = \{\mathbf{x} \in \mathbb{R}^n : h(\mathbf{x}) \leq 0\}.$$

The ball approximation $\mathcal{B}_h(\mathbf{x})$ at $\mathbf{x} \in \mathcal{H}$ is expressed in terms of a center map $\mathbf{c} : \mathcal{H} \to \mathbb{R}^n$ and a radius map $r : \mathcal{H} \to \mathbb{R}$:

$$\mathcal{B}_h(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{y} - \mathbf{c}(\mathbf{x})\| \leq r(\mathbf{x})\}.$$

These two maps must satisfy the following conditions:

B1. Both $\mathbf{c}$ and r are continuous on $\mathcal{H}$.
B2. If $h(\mathbf{x}) < 0$, then $\mathbf{x} \in \text{int } \mathcal{B}_h(\mathbf{x})$, the interior of $\mathcal{B}_h(\mathbf{x})$.
B3. If $h(\mathbf{x}) = 0$, then $\mathbf{x} \in \partial\mathcal{B}_h(\mathbf{x})$, and $c(\mathbf{x}) = \mathbf{x} - \alpha\nabla h(\mathbf{x})^{\mathsf{T}}$ for some fixed $\alpha > 0$.

Maps which satisfy B1, B2, and B3 are the following, assuming h is continuously differentiable:

$$\mathbf{c}(\mathbf{x}) = \mathbf{x} - \alpha\nabla h(\mathbf{x})^{\mathsf{T}}, \quad r(\mathbf{x}) = \alpha\|\nabla h(\mathbf{x})\| - \beta h(\mathbf{x}),$$

where α and β are fixed positive scalars.

Let $\mathbf{c}_i$ and r_i denote center and radius maps associated with g_i, let $\mathcal{B}_i$ be the associated ball given by

$$\mathcal{B}_i(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{y} - \mathbf{c}_i(\mathbf{x})\| \leq r_i(\mathbf{x})\},$$

and define $\mathcal{B}(\mathbf{x}) = \cap_{i=1}^m \mathcal{B}_i(\mathbf{x})$. Our generalization of the algorithm of Lin and Han is the following:

Ball approximation algorithm (BAA)

1. Let $\mathbf{x}_0$ be a feasible point for (C).
2. For $k = 0, 1, \ldots$
   (a) Let $\mathbf{y}_k$ be a solution of the problem
   $$\min \quad f(\mathbf{x}) \quad \text{subject to } \mathbf{x} \in \chi \cap \mathcal{B}(\mathbf{x}_k). \tag{6.1}$$
   (b) Set $\mathbf{x}_{k+1} = \mathbf{x}(\tau_k)$ where $\mathbf{x}(\tau) = (1-\tau)\mathbf{x}_k + \tau\mathbf{y}_k$ and τ_k is the largest $\tau \in [0, 1]$ such that $\mathbf{x}(\sigma) \in \mathcal{F}$ for all $\sigma \in [0, \tau]$.

In [16, Lem. 3.1] it is shown that int $\mathcal{B}(\mathbf{x}) \neq \emptyset$ for each $\mathbf{x} \in \mathcal{F}$ when the center and radius maps $\mathbf{c}_i$ and r_i satisfy B2 and B3 and there exists $\bar{\mathbf{x}}$ such that $\mathbf{g}(\bar{\mathbf{x}}) < \mathbf{0}$. Lin's proof is based on the following observation: For $\epsilon > 0$ sufficiently small, $\mathbf{x} + \epsilon(\bar{\mathbf{x}} - \mathbf{x})$ lies in the interior of $\mathcal{B}_i(\mathbf{x})$ for each i. In C2 we also assume that $\bar{\mathbf{x}} \in \text{ri } \chi$, where "ri" denotes relative interior. Hence, for $\epsilon > 0$ sufficiently small, $\mathbf{x} + \epsilon(\bar{\mathbf{x}} - \mathbf{x})$ lies in both ri χ and in the interior of $\mathcal{B}_i(\mathbf{x})$ for each i. Consequently, we have

$$\text{ri } \chi \cap \text{int } \mathcal{B}(\mathbf{x}) \neq \emptyset \text{ for every } \mathbf{x} \in \mathcal{F}. \tag{6.2}$$

This implies that the subproblems (6.1) of BAA are always feasible. An optimal solution $\mathbf{y}_k$ exists due to the compactness of the feasible set and the continuity of the objective function.

THEOREM 6.1. *If* C1, C2, *and* C3 *hold and the center map* $\mathbf{c}_i$ *and the radius map* r_i *satisfy* B1, B2, *and* B3, $i = 1, 2, \ldots, m$, *then the limit* $\mathbf{x}^*$ *of any convergent subsequence of iterates* $\mathbf{x}_k$ *of Algorithm* 2 *is a solution of* (C).

Proof. Initially, $\mathbf{x}_0 \in \mathcal{F}$. Proceeding by induction, it follows from the line search in Step 2a of BAA that $\mathbf{x}_k \in \mathcal{F}$ for each k. By B2 and B3, $\mathbf{x} \in \mathcal{B}_h(\mathbf{x})$ if $h(\mathbf{x}) \le 0$. Consequently, $\mathbf{x}_k \in \chi \cap \mathcal{B}(\mathbf{x}_k)$ for each k. This shows that $\mathbf{x}_k$ is feasible in (6.1) for each k, and the minimizer $\mathbf{y}_k$ in (6.1) satisfies

$$f(\mathbf{y}_k) \le f(\mathbf{x}_k) \quad \text{for each } k. \tag{6.3}$$

By the convexity of f and by (6.3), we have

$$f(\mathbf{x}_{k+1}) \le \tau_k f(\mathbf{y}_k) + (1 - \tau_k) f(\mathbf{x}_k) \le f(\mathbf{x}_k), \tag{6.4}$$

where $\tau_k \in [0, 1]$ is defined in Step 2b of BAA. Hence, $f(\mathbf{x}_k)$ approaches a limit monotonically. Since $\mathcal{F}$ is compact and $\mathbf{x}_k \in \mathcal{F}$ for each k, an accumulation point $\mathbf{x}^* \in \mathcal{F}$ exists. Since the center maps $\mathbf{c}_i$ and the radius maps r_i are continuous, the balls $\mathcal{B}_i(\mathbf{x}_k)$ are uniformly bounded, and hence, the $\mathbf{y}_k$ are contained in bounded set. Let $\mathbf{y}^*$ denote an accumulation point of the $\mathbf{y}_k$. To simplify the exposition, let $(\mathbf{x}_k, \mathbf{y}_k)$ denote a pruned version of the original sequence which approaches the limit $(\mathbf{x}^*, \mathbf{y}^*)$.

We now show that

$$\mathbf{y}^* = \arg\ \min \quad f(\mathbf{x}) \quad \text{subject to } \mathbf{x} \in \chi \cap \mathcal{B}(\mathbf{x}^*). \tag{6.5}$$

Suppose, to the contrary, that there exists $\tilde{\mathbf{y}} \in \chi \cap \mathcal{B}(\mathbf{x}^*)$ such that $f(\tilde{\mathbf{y}}) < f(\mathbf{y}^*)$. Referring to the discussion before (6.2), choose $\tilde{\mathbf{x}} \in \text{ri } \chi$ with $\tilde{\mathbf{x}} \in \text{int } \mathcal{B}_i(\mathbf{x}^*)$ for each i. Define $\hat{\mathbf{y}} = \tilde{\mathbf{y}} + \epsilon(\tilde{\mathbf{x}} - \tilde{\mathbf{y}})$ where $\epsilon > 0$ is small enough that $\hat{\mathbf{y}} \in \text{ri } \chi$, $\hat{\mathbf{y}} \in \text{int } \mathcal{B}_i(\mathbf{x}^*)$ for each i, and $f(\hat{\mathbf{y}}) < f(\mathbf{y}^*)$. For k sufficiently large, $\hat{\mathbf{y}} \in \chi \cap \mathcal{B}(\mathbf{x}_k)$ due to the continuity of the center and radius maps. Since $f(\mathbf{y}_k)$ approaches $f(\mathbf{y}^*) > f(\tilde{\mathbf{y}})$, we contradict the optimality of $\mathbf{y}_k$ in (6.1). This establishes (6.5).

Again, by B2 and B3, $\mathbf{x}^*$ is feasible in (6.5). Since $\mathbf{y}^*$ is optimal in (6.5), we have $f(\mathbf{y}^*) \le f(\mathbf{x}^*)$. We will show that

$$f(\mathbf{y}^*) = f(\mathbf{x}^*). \tag{6.6}$$

Suppose, to the contrary, that $f(\mathbf{y}^*) < f(\mathbf{x}^*)$. Since $\mathbf{x}^* \in \mathcal{F}$, we conclude that for each i, one of the following two cases can occur:

(i) $g_i(\mathbf{x}^*) = 0$: In this case, it follows from B3 that $\mathbf{x}^* \in \partial \mathcal{B}_i(\mathbf{x}^*)$. Since both $\mathbf{x}^*$ and $\mathbf{y}^* \in \chi \cap \mathcal{B}_i(\mathbf{x}^*)$, we have $[\mathbf{x}^*, \mathbf{y}^*] \in \chi \cap \mathcal{B}_i(\mathbf{x}^*)$. Hence, the vector $\mathbf{y}^* - \mathbf{x}^*$ makes an acute angle with the inward pointing normal at $\mathbf{x}^*$. By B3 the inward pointing normal is a positive multiple of $-\nabla g(\mathbf{x}^*)$; it follows that

$$-\nabla g(\mathbf{x}^*)(\mathbf{y}^* - \mathbf{x}^*) > 0.$$

By a Taylor expansion around $\mathbf{x}^*$, we see that there exist $\sigma_i \in (0, 1)$ such that

$$g_i(\mathbf{x}^* + \sigma(\mathbf{y}^* - \mathbf{x}^*)) < 0 \quad \text{for all } \sigma \in (0, \sigma_i]. \tag{6.7}$$

(ii) $g_i(\mathbf{x}^*) < 0$: In this case, there trivially exists $\sigma_i \in (0,1)$ such that (6.7) holds.
Let σ^* be the minimum of σ_i, $1 \le i \le m$. By the convexity of f, we have

$$f(\mathbf{x}^* + \sigma^*(\mathbf{y}^* - \mathbf{x}^*)) \le f(\mathbf{x}^*) + \sigma^*(f(\mathbf{y}^*) - f(\mathbf{x}^*)) < f(\mathbf{x}^*) \tag{6.8}$$

since $f(\mathbf{y}^*) < f(\mathbf{x}^*)$). Since both $\mathbf{x}_k$ and $\mathbf{y}_k \in \chi \cap \mathcal{B}(\mathbf{x}_k)$, the line segment $[\mathbf{x}_k, \mathbf{y}_k]$ is contained in $\chi \cap \mathcal{B}(\mathbf{x}_k)$. Since $\mathbf{x}_k$ approaches $\mathbf{x}^*$ and $\mathbf{y}_k$ approaches $\mathbf{y}^*$, it follows from (6.7) that

$$\mathbf{x}_k + \sigma^*(\mathbf{y}_k - \mathbf{x}_k) \in \mathcal{F}$$

for k sufficiently large. Again, by the convexity of f, (6.3), and the fact that τ_k is taken as large as possible so that

$$\mathbf{x}_{k+1} = (1 - \tau_k)\mathbf{x}_k + \tau_k \mathbf{y}_k \in \mathcal{F},$$

we have

$$f(\mathbf{x}_{k+1}) \le f(\mathbf{x}_k) + \tau_k(f(\mathbf{y}_k) - f(\mathbf{x}_k)) \le f(\mathbf{x}_k) + \sigma^*(f(\mathbf{y}_k) - f(\mathbf{x}_k)). \tag{6.9}$$

Since $(\mathbf{x}_k, \mathbf{y}_k)$ converges to $(\mathbf{x}^*, \mathbf{y}^*)$, it follows from (6.8) that

$$\lim_{k\to\infty} f(\mathbf{x}_k) + \sigma^*(f(\mathbf{y}_k) - f(\mathbf{x}_k)) = f(\mathbf{x}^*) + \sigma^*(f(\mathbf{y}^*) - f(\mathbf{x}^*)) < f(\mathbf{x}^*). \tag{6.10}$$

Hence, for k sufficiently large, (6.9) and (6.10) imply that $f(\mathbf{x}_{k+1}) < f(\mathbf{x}^*)$, which contradicts the monotone decreasing convergence (6.4) of $f(\mathbf{x}_k)$ to $f(\mathbf{x}^*)$. This completes the proof of (6.6).

Let $L : \mathbb{R}^{m+n} \to \mathbb{R}$ be the Lagrangian defined by

$$L(\boldsymbol{\lambda}, \mathbf{x}) = f(\mathbf{x}) + \frac{1}{2}\sum_{i=1}^{m} \lambda_i \left(\|\mathbf{x} - \mathbf{c}_i(\mathbf{x}^*)\|^2 - r_i(\mathbf{x}^*) \right).$$

Since $\mathbf{y}^*$ is a solution of (6.5) and the Slater condition (6.2) holds, the first-order optimality condition holds at $\mathbf{y}^*$. That is, there exist $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that

$$\left.\begin{array}{l} \boldsymbol{\lambda}^* \ge \mathbf{0}, \quad \lambda_i^*(\|\mathbf{y}^* - \mathbf{c}_i(\mathbf{x}^*)\|^2 - r_i^2(\mathbf{x}^*)) = 0, \;\; i = 1, 2, \ldots, m, \\ \nabla_x L(\boldsymbol{\lambda}^*, \mathbf{y}^*)(\mathbf{x} - \mathbf{y}^*) \ge \mathbf{0} \text{ for all } \mathbf{x} \in \chi. \end{array}\right\} \tag{6.11}$$

If $\nabla f(\mathbf{y}^*)(\mathbf{x} - \mathbf{y}^*) \ge \mathbf{0}$ for all $\mathbf{x} \in \chi$, then $\mathbf{y}^*$ is the global minimizer of the convex function f over χ. Since $f(\mathbf{y}^*) = f(\mathbf{x}^*)$ by (6.6), it follows that $\mathbf{x}^*$ is a solution of (C), and the proof would be complete. Hence, we suppose that $\nabla f(\mathbf{y}^*)(\mathbf{x} - \mathbf{y}^*) < \mathbf{0}$ for some $\mathbf{x} \in \chi$, which implies that $\boldsymbol{\lambda}^* \neq \mathbf{0}$ by (6.11).

Since f is convex, we have

$$f(\mathbf{x}^*) \ge f(\mathbf{y}^*) + \nabla f(\mathbf{y}^*)(\mathbf{x}^* - \mathbf{y}^*). \tag{6.12}$$

We expand the expression

$$\frac{1}{2}\sum_{i=1}^{m} \lambda_i \left(\|\mathbf{x} - \mathbf{c}_i(\mathbf{x}^*)\|^2 - r_i^2(\mathbf{x}^*) \right)$$

in a Taylor series around $\mathbf{x} = \mathbf{y}^*$ and evaluate at $\mathbf{x} = \mathbf{x}^*$ to obtain

$$\frac{1}{2}\sum_{i=1}^{m} \lambda_i \left(\|\mathbf{x}^* - \mathbf{c}_i(\mathbf{x}^*)\|^2 - r_i^2(\mathbf{x}^*)\right) = \frac{1}{2}\sum_{i=1}^{m} \lambda_i \left(\|\mathbf{y}^* - \mathbf{c}_i(\mathbf{x}^*)\|^2 - r_i^2(\mathbf{x}^*)\right)$$
$$+\sum_{i=1}^{m} \lambda_i (\mathbf{y}^* - \mathbf{c}_i(\mathbf{x}^*))^{\mathsf{T}}(\mathbf{x}^* - \mathbf{y}^*) + \frac{1}{2}\|\mathbf{x}^* - \mathbf{y}^*\|^2 \sum_{i=1}^{m} \lambda_i^*.$$

We add this equation to (6.12) to obtain

$$L(\mathbf{x}^*, \boldsymbol{\lambda}^*) \geq L(\mathbf{y}^*, \boldsymbol{\lambda}^*) + \nabla_x L(\mathbf{y}^*, \boldsymbol{\lambda}^*)(\mathbf{x}^* - \mathbf{y}^*) + \frac{1}{2}\|\mathbf{x}^* - \mathbf{y}^*\|^2 \sum_{i=1}^{m} \lambda_i^*. \tag{6.13}$$

By complementary slackness and by (6.6), we have $L(\mathbf{y}^*, \boldsymbol{\lambda}^*) = f(\mathbf{y}^*) = f(\mathbf{x}^*)$. Hence, (6.13) yields

$$\frac{1}{2}\|\mathbf{x}^* - \mathbf{y}^*\|^2 \sum_{i=1}^{m} \lambda_i^* \leq -\nabla_x L(\mathbf{y}^*, \boldsymbol{\lambda}^*)(\mathbf{x}^* - \mathbf{y}^*)$$
$$+\frac{1}{2}\sum_{i=1}^{m} \lambda_i^* \left(\|\mathbf{x}^* - \mathbf{c}_i(\mathbf{x}^*)\|^2 - r_i^2(\mathbf{x}^*)\right). \tag{6.14}$$

By (6.11) and the fact that $\mathbf{x}^* \in \chi$, we have $\nabla_x L(\mathbf{y}^*, \boldsymbol{\lambda}^*)(\mathbf{x}^* - \mathbf{y}^*) \geq \mathbf{0}$. Since $\mathbf{x}^* \in \mathcal{B}(\mathbf{x}^*)$, the last term in (6.14) is nonpositive. Hence, the entire right side of (6.14) is nonpositive. Since $\boldsymbol{\lambda}^* \geq \mathbf{0}$ and $\boldsymbol{\lambda}^* \neq \mathbf{0}$, (6.14) implies that $\mathbf{y}^* = \mathbf{x}^*$.

Replacing $\mathbf{y}^*$ by $\mathbf{x}^*$ in the first-order conditions (6.11) gives

$$\left(\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* (\mathbf{x}^* - \mathbf{c}_i(\mathbf{x}^*))^{\mathsf{T}}\right)(\mathbf{x} - \mathbf{x}^*) \geq \mathbf{0} \text{ for all } \mathbf{x} \in \chi. \tag{6.15}$$

If $g_i(\mathbf{x}^*) < 0$, then by B2, $\mathbf{x}^* \in \text{ int } \mathcal{B}_i(\mathbf{x}^*)$ and $\lambda_i^* = 0$ by complementary slackness. If $g_i(\mathbf{x}^*) = 0$, then by B3, $\mathbf{c}_i(\mathbf{x}^*) = \mathbf{x}^* - \alpha \nabla g_i(\mathbf{x}^*)^{\mathsf{T}}$. With these substitutions, (6.15) yields

$$\left(\nabla f(\mathbf{x}^*) + \alpha \sum_{i=1}^{m} \lambda_i^* \nabla g_i(\mathbf{x}^*)\right)(\mathbf{x} - \mathbf{x}^*) \geq \mathbf{0} \text{ for all } \mathbf{x} \in \chi.$$

Hence, the first-order optimality conditions for (C) are satisfied at $\mathbf{x}^*$. Since the objective function and the constraints of (C) are convex, $\mathbf{x}^*$ is a solution of (C). This completes the proof. □

7. Numerical experiments. We investigate the performance of the algorithms of the previous sections using randomly generated quadratically constrained quadratic programming problems of the form

$$\min \mathbf{x}^{\mathsf{T}}\mathbf{A}_0\mathbf{x} + \mathbf{b}_0^{\mathsf{T}}\mathbf{x} \text{ subject to } \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \tag{QP}$$

where $\mathbf{x} \in \mathbb{R}^n$ and $g_i(\mathbf{x}) = \mathbf{x}^{\mathsf{T}}\mathbf{A}_i\mathbf{x} + \mathbf{b}_i^{\mathsf{T}}\mathbf{x} + c_i$, $i = 1, 2, \ldots, m$. Here $\mathbf{b}_i \in \mathbb{R}^n$ and c_i is a scalar for each i. The matrices $\mathbf{A}_i$ are symmetric, positive definite for $i \geq 1$. In our experiments with the ball approximation algorithm, we take $\mathbf{A}_0$ symmetric, positive semidefinite. In our experiments with the branch and bound algorithm, we consider more general indefinite $\mathbf{A}_0$. The codes are written in either C or Fortran. The experiments were implemented using a Matlab 7.0.1 interface on a PC with 2GB memory and Intel Core 2 Duo 2Ghz processors running the Windows Vista operating system.

7.1. Rate of convergence for BAA. The theory of Section 6 establishes the convergence of BAA. Experimentally, we observe that the convergence rate is linear. Figure 7.1 shows that the behavior of the KKT error as a function of the iteration number for a randomly generated positive definite matrix $\mathbf{A}_0$ of dimension 200 and for 4 ellipsoidal constraints ($m = 4$). The KKT error is computed using the formula given

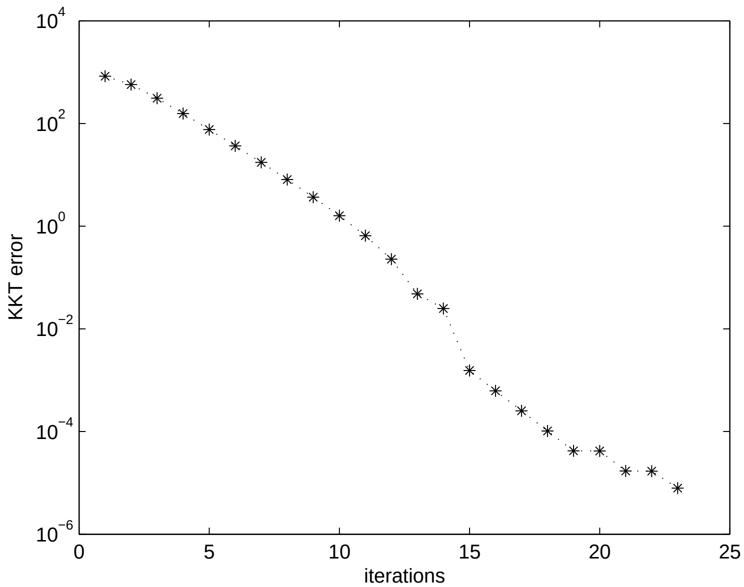

FIG. 7.1. *KKT error versus iteration number for $n = 200$, $m = 4$, and $\mathbf{A}_0$ positive definite*

in Section 4 of [10]. Roughly, this formula amounts to the infinity norm of the gradient of the Lagrangian plus the infinity norm of the violation in complementary slackness. If $\mathbf{A}_0$ is constructed to have precisely one zero eigenvalue, then the convergence rate again appears to be linear, as seen in Figure 7.2.

7.2. Comparison with other algorithms for programs with convex cost. To gain some insight into the relative performance of the ball approximation algorithm (BAA), we solved randomly generated problems with convex cost using three other algorithms:

- SEDUMI, for optimization over symmetric cones.
- The gradient projection algorithm. We tried both the nonmonotone gradient project algorithm (NGPA) given in [11] and the nonmonotone spectral projected gradient method (SPG) of Birgin, Martínez, and Raydan [2, 3] (ACM Algorithm 813).

We now discuss in detail how each of these algorithms was implemented. The BAA subproblems (6.1) have the form

$$\min \mathbf{x}^{\mathsf{T}}\mathbf{A}_0\mathbf{x} + \mathbf{b}_0^{\mathsf{T}}\mathbf{x} \quad \text{subject to } \|\mathbf{x} - \mathbf{c}_i\|^2 \le r_i^2, \quad 1 \le i \le m. \tag{7.1}$$

We solve these subproblems by applying the active set algorithm (ASA) developed in [11] to the dual problem. To facilitate the evaluation of the dual function, we compute the diagonalization $\mathbf{A}_0 = \mathbf{QDQ}^{\mathsf{T}}$ where $\mathbf{D}$ is diagonal and $\mathbf{Q}$ is orthogonal.

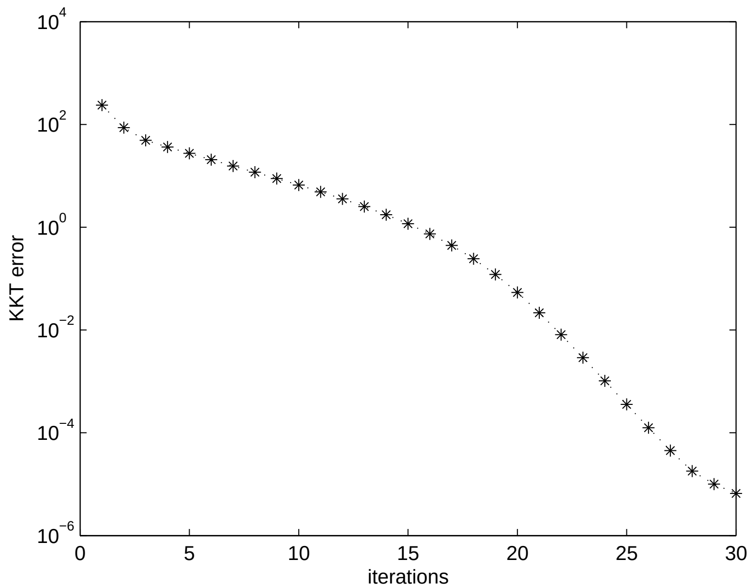

FIG. 7.2. *KKT error versus iteration number for $n = 200$, $m = 4$, and a positive semidefinite $\mathbf{A}_0$.*

Substituting $\mathbf{x} = \mathbf{Q}\mathbf{y}$ in (7.1) yields the equivalent problem

$$\min \mathbf{y}^\mathsf{T}\mathbf{D}\mathbf{y} + \mathbf{b}_0^\mathsf{T}\mathbf{Q}\mathbf{y} \quad \text{subject to } \|\mathbf{y} - \mathbf{Q}^\mathsf{T}\mathbf{c}_i\|^2 \le r_i^2, \quad 1 \le i \le m.$$

The dual problem is

$$\max_{\boldsymbol{\lambda} \ge \mathbf{0}} \ \min_{\mathbf{y} \in \mathbb{R}^n} \ \mathbf{y}^\mathsf{T}\mathbf{D}\mathbf{y} + \mathbf{b}_0^\mathsf{T}\mathbf{Q}\mathbf{y} + \sum_{i=1}^m \lambda_i \left(\|\mathbf{y} - \mathbf{Q}^\mathsf{T}\mathbf{c}_i\|^2 - r_i^2 \right). \tag{7.2}$$

The i-th component of the gradient of the dual function with respect to $\boldsymbol{\lambda}$ is simply $\|\mathbf{y}(\boldsymbol{\lambda}) - \mathbf{Q}^\mathsf{T}\mathbf{c}_i\|^2 - r_i^2$ where $\mathbf{y}(\boldsymbol{\lambda})$ achieves the minimum in (7.2). This minimum is easily evaluated since the quadratic term in the objective function is diagonal.

SEDUMI could be applied directly to (QP) when the cost function is strongly convex. We used Version 1.1 of the code obtained from

http://sedumi.mcmaster.ca/

In implementing the gradient projection algorithm for (QP), we need to project a vector onto the feasible set. This amounts to solving a problem of the form

$$\min \|\mathbf{x} - \mathbf{a}\|^2 \text{ subject to } \mathbf{g}(\mathbf{x}) \le \mathbf{0}.$$

We solved this problem using BAA. An iteration of BAA reduces to the solution of a problem with the following structure:

$$\min \|\mathbf{x} - \mathbf{a}\|^2 \text{ subject to } \|\mathbf{x} - \mathbf{c}_i\|^2 \le r_i^2, \quad i = 1, 2, \ldots, m. \tag{7.3}$$

As in [16], we solve these problems by forming the dual problem

$$\max_{\boldsymbol{\lambda} \ge \mathbf{0}} \ \min_{\mathbf{x} \in \mathbb{R}^n} \ \|\mathbf{x} - \mathbf{a}\|^2 + \sum_{i=1}^m \lambda_i \left(\|\mathbf{x} - \mathbf{c}_i\|^2 - r_i^2 \right).$$

Table 7.1
Positive definite cases

n, m		SED	success	BAA	success	NGPA	success	SPG	success
	time	0.52		0.07		4.70		5.94	
100,4	*iter*	10.06	28	19.00	30	172.43	17	200.06	17
	time	2.75		0.32		10.68		11.76	
200,4	*iter*	9.56	26	12.70	30	203.23	21	348.36	20
	time	8.14		0.72		128.19		122.23	
300,4	*iter*	9.60	27	21.46	30	269.86	20	431.83	20
	time	20.13		1.64		404.28		438.84	
400,4	*iter*	10.26	27	49.26	29	352.66	18	545.23	18
	time	44.07		2.54		579.21		647.56	
500,4	*iter*	12.33	28	29.90	30	369.20	15	574.13	13
	time	57.28		4.27		648.79		611.60	
600,4	*iter*	9.30	26	36.80	29	309.43	19	306.50	19
	time	3.51		0.08		86.89		81.67	
100,40	*iter*	10.26	28	19.00	30	150.76	21	165.66	21
	time	26.56		0.32		268.63		250.22	
200,40	*iter*	12.70	30	12.70	30	199.70	17	218.50	16
	time	54.56		0.81		732.50		727.92	
200,100	*iter*	10.66	30	9.50	30	295.80	20	327.26	20
	time	23.84		0.72		579.62		530.02	
100,200	*iter*	14.96	30	20.06	30	249.43	18	261.46	19
	time	0.093		0.002		3.02		2.75	
4,100	*iter*	9.06	29	6.96	30	19.46	26	19.40	25
	time	0.114		0.004		6.23		5.70	
4,200	*iter*	9.56	27	8.73	30	16.26	26	16.46	26
	time	0.148		0.012		13.45		11.58	
4,300	*iter*	11.06	25	12.56	30	15.26	24	15.33	23
	time	0.195		0.014		16.27		12.87	
4,400	*iter*	13.33	28	12.26	30	16.26	28	15.70	28
	time	0.221		0.017		21.08		18.16	
4,500	*iter*	13.83	26	11.50	30	13.83	26	13.83	26
	time	0.235		0.018		31.65		34.83	
4,600	*iter*	12.13	26	11.00	30	15.40	24	16.33	24

After carrying out the inner minimization, this reduces to

$$\max_{\boldsymbol{\lambda}\geq\mathbf{0}} -\frac{\|\mathbf{a}+\sum_{i=1}^m \lambda_i \mathbf{c}_i\|^2}{1+\sum_{i=1}^m \lambda_i} + \sum_{i=1}^m \lambda_i(\|\mathbf{c}_i\|^2 - r_i^2). \tag{7.4}$$

If $\boldsymbol{\lambda}$ solves the dual problem (7.4), then the associated solution of the primal problem (7.3) is

$$\mathbf{x} = \frac{\mathbf{a}+\sum_{i=1}^m \lambda_i \mathbf{c}_i}{1+\sum_{i=1}^m \lambda_i}.$$

Again, the dual problem (7.4) is solved using the active set algorithm (ASA) of [11].

The test problems used in Tables 7.1 and 7.2 were generated as follows: Let $\mathrm{Rand}(n, l, u)$ denote a vector in $\mathbb{R}^n$ whose entries are chosen randomly in the interval

(l, u). Random positive definite matrices $\mathbf{A}$ are generated using the procedure given in [17], which we now summarize. Let $\mathbf{w}_i \in \text{Rand}(n, -1, 1)$ for $i = 1, 2, 3$, and define

$$\mathbf{Q}_i = \mathbf{I} - 2\mathbf{v}_i\mathbf{v}_i^{\mathsf{T}}, \quad \mathbf{v}_i = \mathbf{w}_i/\|\mathbf{w}_i\|.$$

Let $\mathbf{D}$ be a diagonal matrix with diagonal in $\text{Rand}(n, 0, 100)$. Finally, $\mathbf{A} = \mathbf{UDU}^{\mathsf{T}}$ with $\mathbf{U} = \mathbf{Q}_1\mathbf{Q}_2\mathbf{Q}_3$. To obtain a randomly generated positive semidefinite matrix, we use the same procedure, however, we randomly set one diagonal element of $\mathbf{D}$ to zero.

We make a special choice for c_i to ensure that the feasible set for (QP) is nonempty. We first generate $\mathbf{p} \in \text{Rand}(n, -50, 50)$ and we set

$$c_i = -(\mathbf{p}^{\mathsf{T}}\mathbf{A}_i\mathbf{p} + \mathbf{b}_i^{\mathsf{T}}\mathbf{p} + s_i),$$

where s_i is randomly generated in the interval $[0, 10]$ and $\mathbf{b}_i \in \text{Rand}(n, -100, 100)$. With this choice for c_i, the feasible set for (QP) is nonempty since $\mathbf{p}$ lies in the interior of the feasible set. The stopping criterion in our experiments was

$$\|P(\mathbf{x}_k - \mathbf{g}_k) - \mathbf{x}_k\| \le 10^{-4}, \tag{7.5}$$

where P denotes projection into the feasible set for (QP) and $\mathbf{g}_k = 2\mathbf{A}_0\mathbf{x}_k + \mathbf{b}_0$ is the gradient of the objective function at $\mathbf{x}_k$. When the cost is convex, the left side of (7.5) vanishes if and only if $\mathbf{x}_k$ is a solution of (QP).

Tables 7.1 and 7.2 report the average CPU time in seconds (*time*), the average number of iterations (*iter*), and the number of successes in 30 randomly generated test problems. The algorithm was considered successful if the error tolerance (7.5) was satisfied.

Based on our numerical experiments, it appears that BAA can achieve an error tolerance on the order of the square root of the machine epsilon [9, 24], similar to the computing precision which is achieved by interior point methods for linear programming prior to simplex crossover. The convergence tolerance (7.5) was chosen since it seems to approach the maximum accuracy which could be achieved by BAA in these test problems. Numerically, BAA seems to terminate when the solution to the subproblem (6.1) yields a direction which departs from the feasible set, and hence, the stepsize in the line search Step 2b is zero. We were able to achieve a further improvement in the solution by taking a partial step in this infeasible direction since the increase in constraint violation was much less than the improvement in objective function value. Nonetheless, the improvement in accuracy achieved by permitting infeasibility was at most one digit in our experiments.

In Tables 7.1 and 7.2 we see that BAA gave the best results for this test set, both in terms of CPU time and in terms of successes (the number of times that the convergence tolerance (7.5) was achieved). Recall that the gradient projection algorithms in our experiments used BAA to compute the projected gradient. The convergence failures for the gradient projection algorithms in Tables 7.1 and 7.2 were due to the fact that BAA was unable to compute the projected gradient with enough accuracy to yield descent in the gradient projection algorithm.

7.3. Problems with nonconvex cost. We tested our ellipsoidal branch and bound algorithm using some randomly generated test problems with $\mathbf{A}_0$ indefinite. To compute $\boldsymbol{\mu}$ in (3.4), we used the power method (see [24]) to find the eigenvector associated with the largest eigenvalue. We chose λ in (1.2) to be 0.1 minus the smallest eigenvalue of $\mathbf{A}_0$. NGPA was used to locally solve (QP) and update the upper bound.

TABLE 7.2
Positive semidefinite cases

n, m		BAA	success	NGPA	success	SPG	success
	time	0.11		15.04		17.38	
100,4	*iter*	42.16	30	328.53	22	408.20	19
	time	0.55		44.32		45.82	
200,4	*iter*	99.33	30	313.10	20	356.43	22
	time	1.02		290.19		304.89	
300,4	*iter*	74.23	30	374.13	22	417.60	21
	time	2.28		501.14		572.63	
400,4	*iter*	111.83	30	404.66	19	492.83	19
	time	5.37		620.13		657.61	
500,4	*iter*	200.03	27	382.66	16	478.60	17
	time	0.61		356.40		321.92	
100,40	*iter*	82.30	30	276.56	22	237.23	21
	time	2.74		398.54		415.28	
200,40	*iter*	127.23	30	369.43	17	416.060	17
	time	3.19		1030.02		949.29	
100,200	*iter*	108.63	28	311.40	19	352.23	18
	time	0.054		16.74		14.25	
4,100	*iter*	38.66	30	31.13	16	31.23	16
	time	0.075		44.74		33.24	
4,200	*iter*	43.46	27	26.20	13	23.36	18
	time	0.076		111.03		100.38	
4,300	*iter*	31.50	29	29.33	14	28.60	11
	time	0.049		205.17		237.62	
4,400	*iter*	36.73	29	27.20	18	31.23	18
	time	0.065		229.77		247.82	
4,500	*iter*	41.86	28	24.46	16	26.30	17

We took $m = 2$ and randomly generated test problem using the procedure in [1]. That is, the ellipsoidal constraint functions in (QP) have the form

$$g_i(\mathbf{x}) = (\mathbf{x} - \mathbf{c}_i)^{\mathsf{T}} \mathbf{B}_i^{-1} (\mathbf{x} - \mathbf{c}_i) - 1,$$

where $\mathbf{B}_i = \mathbf{U}\mathbf{D}_i\mathbf{U}^{\mathsf{T}}$ and $\mathbf{U}$ is as given earlier. $\mathbf{D}_i$ is a diagonal matrix with its diagonal in $\mathrm{Rand}(n, 0, 60)$, $\mathbf{c}_1 \in \mathrm{Rand}(n, 0, 100)$, and $\mathbf{c}_2 = \mathbf{c}_1 + .8\mathbf{v}$ where $\mathbf{v}$ is the semi-major axis of the ellipsoid $g_1(\mathbf{x}) \le 0$. For this choice of $\mathbf{c}_2$, the ellipsoids $g_1(\mathbf{x}) \le 0$ and $g_2(\mathbf{x}) \le 0$ have nonempty intersection at $\mathbf{x} = \mathbf{c}_2$. In the objective function, $\mathbf{A}_0 = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathsf{T}}$ where $\mathbf{D}$ is a diagonal matrix with diagonal in $\mathrm{Rand}(n, -30, 30)$ and $\mathbf{b}_0 \in \mathrm{Rand}(n, -1, 1)$. The case $m = 2$ is especially important since quadratic problems with two ellipsoidal constraints belong to the class of Celis-Dennis-Tapia subproblems [4] which arise from the application of the trust region method for equality constrained nonlinear programming [12, 20, 6, 5, 15, 26, 19].

If UB_k and LB_k are the respective upper and lower bounds for the optimal objective function value at iteration k, then our stopping criterion was

$$\mathrm{UB}_k - \mathrm{LB}_k \le \max\{\epsilon_a, \epsilon_r |\mathrm{LB}_k|\},$$

with $\epsilon_a = 10^{-5}$ and $\epsilon_r = 10^{-2}$.

We considered problems of 8 different dimensions ranging from 30 up to 300 as shown in Table 7.3. For each dimension, we solved 4 randomly generated problems.

Table 7.3 shows the numerical results for our test instances, where "*neigs*" is the number of negative eigenvalues of the objective function, "lb^1" and "ub^1" are the lower bound and upper bounds at the first step, "*val*" is the computed optimal value and "*it*" is the number of iterations. We also report the performance of the algorithm for $m = 6$ in Table 7.4.

TABLE 7.3
The performance of branch and bound algorithm for $m = 2$

n	*neigs*	lb^1	ub^1	*val*	*it*	*time*
30	12	34827.3	35256.3	35254.8	5	1.75
	17	-41212.1	-40746.2	-40748.8	21	3.58
	14	38601.4	38977.6	38977.6	0	0.72
	17	-31534.2	-31108.4	-31119.8	92	8.82
50	22	-357168.8	-356828.8	-356828.8	0	0.52
	21	-33792.9	-33447.9	-33447.9	1	0.84
	21	-29694.6	-29254.1	-29255.2	247	23.08
	26	35034.0	35416.6	35414.8	5	2.12
60	29	17783.6	18227.4	18227.4	78	22.41
	26	-27498.2	-27110.5	-27110.5	69	18.22
	30	-69845.7	-69463.1	-69463.1	0	0.56
	28	20408.7	20963.1	20927.1	273	42.65
100	50	-11495.2	-11196.9	-11218.9	56	30.72
	51	17539.6	17909.3	17909.3	4	1.84
	52	-46065.5	-45653.2	-45653.2	0	0.88
	40	970829.8	971326.6	971326.6	0	0.92
150	75	-302382.2	-302071.0	-302071.0	0	0.95
	83	29089.4	29500.8	29500.8	64	31.45
	72	16580.5	16904.9	16904.9	1	1.98
	73	-32461.9	-32036.1	-32036.1	1	1.37
200	100	10798.5	11226.1	11226.1	81	56.58
	95	-27242.9	-26792.1	-26792.1	2	2.27
	100	35293.0	35862.1	35862.1	1	1.63
	96	-31712.8	-31138.3	-31138.3	77	47.06
250	135	37015.8	37477.6	37477.6	1	2.9
	131	-27278.9	-26563.0	-26780.0	86	88.40
	128	-9979.6	-9683.9	-9683.9	59	131.54
	121	-371385.9	-370991.9	-370991.9	0	2.03
300	145	-162041.5	-161645.7	-161645.7	0	5.33
	152	-48085.4	-47529.3	-47529.3	1	7.56
	138	226345.6	226377.8	226377.8	0	4.79
	148	-17649.5	-17013.7	-17323.2	109	257.52

In comparing our ellipsoidal branch and bound algorithm based on linear underestimation (EBL) to the ellipsoidal branch and bound algorithm of Le Thi Hoai An [1] based on dual underestimation (EBD), an advantage of EBD is that the underestimates are often quite tight in the dual-based approach. As seen in Table 7.3, EBL required up to 273 bisections for this test set while EBD in [1] was able to solve randomly generated test problems without any bisections. On the other hand, a disadvantage of EBD is that the dual problems are nondifferentiable when $\mathbf{A}_0$ is indefinite. Consequently, the evaluation of the lower bound using EBD entails solving an optimization problem which, in general, is nondifferentiable. With EBL, how-

ever, computing a lower bound involves solving a convex optimization problem. To summarize, EBD provides tight lower bounds using a nondifferentiable optimization problem for the lower bound, while EBL provides less tight lower bounds using a convex optimization problem for the lower bound.

Table 7.4
The performance of branch and bound algorithm for $m = 6$

n	$neigs$	lb^1	ub^1	val	it	$time$
30	17	22717.1	22993.1	22993.1	1	2.41
	21	-22847.0	-22520.4	-22524.5	14	10.58
	20	-17858.2	-17573.2	-17573.2	1	1.84
60	33	-21818.1	-21489.1	-21489.2	33	21.64
	27	47683.8	47826.0	47826.0	0	2.82
	31	-4926.5	-4652.0	-4728.7	4	7.62
100	56	-35438.9	-35411.0	-35411.0	0	0.78
	52	-1740.1	-1187.5	-1198.2	354	283.25
	49	-6756.5	-6148.9	-6148.9	3	8.06

8. Conclusions. A globally convergent branch and bound algorithm was developed in which the objective function was written as the difference of convex functions. The algorithm was based on an affine underestimate given in Theorem 3.1 for the concave part of the objective function restricted to an ellipsoid. An algorithm of Lin and Han [16, 17] for projecting a point onto a convex set was generalized so as to replace their norm objective by an arbitrary convex function. This generalization could be employed in the branch and bound algorithm for a general objective function when the constraints are convex. Numerical experiments were given for a randomly generated quadratic objective function and randomly generated convex, quadratic constraints.

REFERENCES

[1] L. T. H. An, *An efficient algorithm for globally minimizing a quadratic function under convex quadratic constraints*, Math. Program., 87 (2000), pp. 401–426.

[2] E. G. Birgin, J. M. Martínez, and M. Raydan, *Nonmonotone spectral projected gradient methods for convex sets*, SIAM J. Optim., 10 (2000), pp. 1196–1211.

[3] ———, *Algorithm 813: SPG - software for convex-constrained optimization*, ACM Trans. Math. Softw., 27 (2001), pp. 340–349.

[4] M. Celis, J. E. Dennis, and R. A. Tapia, *A trust region strategy for nonlinear equality constrained optimization*, in Numerical Optimization 1984, Philadelphia, PA, 1985, SIAM, pp. 71–82.

[5] X. D. Chen and Y. Yuan, *A note on quadratic forms*, Math. Program., 86 (1999), pp. 187–197.

[6] ———, *On local solutions of the cdt subproblem*, SIAM J. Optim., 10 (1999), pp. 359–383.

[7] P. L. Combettes and T. Pennanen, *Proximal methods for cohypomonotone otperators*, SIAM J. Control, 43 (2004), pp. 731–742.

[8] C. A. Floudas and V. Visweswaran, *Quadratic optimization*, in Handbook of Global Optimization, R.Horst and P. Pardalos, eds., Kluwer Academic, 1994, pp. 217–270.

[9] W. W. Hager, *Applied Numerical Linear Algebra*, Prentice-Hall, Englewood Cliffs, NJ, 1988.

[10] W. W. Hager and S. Gowda, *Stability in the presence of degeneracy and error estimation*, Math. Program., 85 (1999), pp. 181–192.

[11] W. W. Hager and H. Zhang, *A new active set algorithm for box constrained optimization*, SIAM J. Optim., 17 (2006), pp. 526–557.

[12] M. Heinkenschloss, *On the solution of a two ball trust region subproblem*, Math. Programming, 64 (1994), pp. 249–276.

[13] R. Horst, P. M. Pardalos, and N. V. Thoai, *Introduction to Global Optimization*, Kluwer Academic Publishers, Dordrecht, Holland, 1995.

[14] A. N. Iusem, T. Pennanen, and B. F. Svaiter, *Inexact variants of the proximal point algorithm without monotonicity*, SIAM J. Optim., 13 (2003), pp. 1080–1097.
[15] G. D. Li and Y. Yuan, *Computing a celis-dennis-tapia step*, J. Comput. Math., 23 (2005), pp. 463–478.
[16] A. Lin, *A class of method for projection on a convex set*, Advanced Modeling and Optimization, 5 (2003), pp. 211–221.
[17] A. Lin and S. P. Han, *A class of methods for projection on the intersection of several ellipsoids*, SIAM J. Optim., 15 (2005), pp. 129–138.
[18] J. Linderoth, *A simplicial branch and bound algorithm for solving quadratically constrained quadratic programs*, Math. Program., 103 (2005), pp. 251–282.
[19] J. M. Martinez and S. A. Santos, *A trust-region strategy for minimization on arbitrary domains*, Math. Program., 68 (1995), pp. 267–301.
[20] J. M. Peng and Y. Yuan, *Optimality conditions for the minimization of a quadratic with two quadratic constraints*, SIAM J. Optim., 7 (1997), pp. 579–594.
[21] T. Pennanen, *Local convergence of the proximal point algorithm and multiplier methods without monotonicity*, Math. Oper. Res., 27 (2002), pp. 170–191.
[22] U. Raber, *A simplicial branch and bound method for solving nonconvex all-quadratic programs*, J. Global Optim., 13 (1998), pp. 417–432.
[23] N. Z. Shor, *Cut-off method with space extension in convex programming problems*, Cybernetics and System Analysis, 1 (1977), pp. 94–97.
[24] L. N. Trefethen and D. Bau III, *Numerical Linear Algebra*, SIAM, Philadelphia, 1997.
[25] J. P. Vial, *Strong and weak convexity of sets and functions*, Math. Oper. Res., 8(2) (1983), pp. 231–259.
[26] Y. Ye and S. Zhang, *New results on quadratic minimization*, SIAM J. Optim., 14 (2003), pp. 245–267.
[27] D. B. Yudin and A. S. Nemirovski, *Informational complexity and effective methods for the solution of convex extremal problems*, Ekonom. Mat. Metody., 12 (1976), pp. 550–559 (in Russian).